(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 March 2004 (04.03.2004)

PCT

(10) International Publication Number
WO 2004/018621 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US2003/021598

(22) International Filing Date: 9 July 2003 (09.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/395,763	10 July 2002 (10.07.2002)	US
60/417,511	9 October 2002 (09.10.2002)	US
60/444,933	3 February 2003 (03.02.2003)	US

(71) Applicant *(for all designated States except US)*: **THE PENN STATE RESEARCH FOUNDATION** [US/US]; Intellectual Property Office, 113 Technology Center, 200 Innovation Blvd., University Park, PA 16802 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MARANAS, Costas, D.** [US/US]; Penn State Research Foundation, 304 Old Main, University Park, PA 16802-1504 (US). **BURGARD, Anthony, R.** [US/US]; Penn State Research Foundation, 304 Old Main, University Park, PA 16802-1504 (US). **PHARKYA, Priti** [IN/US]; Penn State University, 304 Old Main, University Park, PA 16802-1504 (US).

(74) Agent: **GOODHUE, John, D.**; McKee, Voorhees & Sease PLC, 801 Grand Avenue, Suite 3200, Des Moines, IA 50309 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

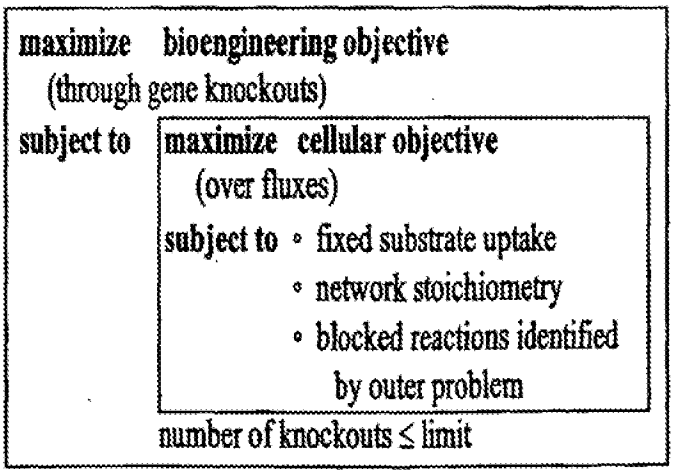
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR DETERMINING GENE KNOCKOUT STRATEGIES



(57) Abstract: A method for determining candidates for gene deletions and additions using a model of a metabolic network associated with an organism, the model includes a plurality of metabolic reactions defining metabolite relationships, the method includes selecting a bioengineering objective for the organism, selecting at least one cellular objective, forming an optimization problem that couples the at least one cellular objective with the bioengineering objective, and solving the optimization problem to yield at least one candidate.

WO 2004/018621 A2

TITLE: METHOD FOR DETERMINING GENE KNOCKOUT STRATEGIES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is based on U.S. Patent Application Serial No. 60/395,763, filed July 10, 2002; U.S. Patent Application Serial No. 60/417,511, filed October, 9, 2002, and U.S. Patent Application Serial No. 60/444,933, filed February 3, 2003, each of which is herein incorporated by reference in its entirety.

GRANT REFERENCE

This work has been supported by Department of Energy pursuant to Grant No. 58855 and the National Science Foundation Grant No. BES0120277. Accordingly, the U.S. government may have certain rights in the invention.

BACKGROUND OF THE INVENTION

The systematic development of engineered microbial strains for optimizing the production of chemicals or biochemicals is an overarching challenge in biotechnology (Stephanopoulos et al., 1998). However, in the absence of metabolic and genetic engineering interventions, the product yields of many microorganisms are often far below their theoretical maximums. This is expected because cellular metabolism is primed, through natural selection, for the maximum responsiveness to the history of selective pressures rather than for the overproduction of specific chemical compounds. Not surprisingly, the behavior of metabolic networks is governed by internal cellular objectives which are often in direct competition with chemical overproduction targets.

The recent explosion of annotated sequence information along with a wealth of chemical literature has enabled the reconstruction of genome-scale metabolic networks for many microorganisms (Edwards and Palsson, 2000; Schilling and Palsson, 2000; Schilling et al., 2002; Forster et al., 2003). This information, used in the context of the flux balance analysis (FBA) modeling framework (Varma and Palsson, 1993), has been employed extensively to explore the integrated functions of metabolic networks (Burgard and Maranas, 2001; Burgard et al., 2001; Papin et al., 2003; Price et al., 2003). FBA models typically invoke the optimization of a particular cellular objective (e.g., ATP production

(Majewski and Domach, 1990; Ramakrishna et al., 2001), biomass formation (Varma and Palsson, 1993,1994), minimization of metabolic adjustment (Segre et al., 2002)), subject to network stoichiometry, to suggest a likely flux distribution. Stoichiometric models of *Escherichia coli* (*E. coli*) metabolism utilizing the biomass maximization hypothesis have been in some cases successful at (i) predicting the lethality of gene knockouts (Edwards and Palsson, 2000; Badarinarayana et al., 2001), (ii) identifying the correct sequence of byproduct secretion under increasingly anaerobic conditions (Varma et al., 1993), and (iii) quantitatively predicting cellular growth rates under certain conditions (Edwards et al., 2001). Interestingly, recent work suggests that even when FBA predictions under the biomass maximization assumption seem to fail, metabolic networks can be evolved, for certain cases, towards maximum growth (i.e., biomass yield) through adaptive evolution (Ibarra et al., 2002).

The ability to investigate the metabolism of single-cellular organisms at a genomic scale, and thus systemic level, motivates the need for novel computational methods aimed at identifying strain engineering strategies.

Thus, one object, feature, or advantage of the present invention is to provide a method for computationally suggesting the manner in which to achieve bioengineering objectives.

A further object, feature or advantage of the present invention is to determine candidates for gene deletion or addition through use of a model of a metabolic network.

A still further object, feature or advantage of the present invention is to provide an optimized method for computationally achieving a bioengineering objective.

Yet another object, feature or advantage of the present invention is to provide an optimized method for computationally achieving a bioengineering objective that is robust.

One or more of these and/or other objects, features and advantages of the present invention will become apparent after review of the following detailed description of the disclosed embodiments and the appended claims.

BRIEF SUMMARY OF THE INVENTION

The systematic development of engineered microbial strains for optimizing the production of chemical or biochemicals is an overarching challenge in biotechnology. The

advent of genome-scale models of metabolism has laid the foundation for the development of computational procedures for suggesting genetic manipulations that lead to overproduction. The present invention describes a computational framework for suggesting gene deletion strategies leading to the overproduction of chemicals or biochemicals in *E.coli*. This is accomplished by ensuring that a drain towards growth resources (i.e., carbon, redox potential, and energy) is accompanied, due to stoichiometry, by the production of a desired production. Specifically, the computation framework identifies multiple gene deletion combinations that maximally couple a postulated cellular objective (e.g., biomass formation) with externally imposed chemical production targets. This nested structure gives rise to a bilevel optimization problems which are solved based on a transformation inspired by duality theory. This procedure of this framework, by coupling biomass formation with chemical production, suggest a growth selection/adaption system for indirectly evolving overproducing mutants.

One embodiment of the invention is directed to a method for determining candidates for gene deletions and additions. The method uses a model of a metabolic network associated with an organism. The model includes a number of metabolic reactions defining metabolite relationships. The method includes selecting a bioengineering objective for the selecting the organism. Next, at least one cellular objective is selected. An optimization problem is formed that couples the cellular objective with the bioengineering objective. Finally, the optimization problem is solved to yield at least one candidate.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts the bilevel optimization structure of Optknock. The inner problem performs the flux allocation based on the optimization of a particular cellular objective (e.g., maximization of biomass yield, MOMA). The outer problem then maximizes the bioengineering objective (e.g., chemical production) by restricting access to key reactions available to the optimization of the inner problem.

Figure 2 depicts the flux distributions of the (A) wild-type *E. coli*, (B) succinate mutant B, (C) succinate mutant C, and (D) lactate mutant C networks that maximize biomass yield under anaerobic conditions.

Figure 3 shows (A) succinate or (B) lactate production limits under anaerobic conditions for mutant A (dash/dotted line), mutant B (dotted line), mutant C (short dashes) and the wild-type *E. coli* network (solid). The production limits are obtained by separately maximizing and minimizing succinate or lactate production for the biomass yields available to each network. The yellow points depict the solution identified by OptKnock (i.e., maximum chemical production at the maximum biomass yield).

Figure 4 shows the aerobic flux distributions of the (A) wild-type *E. coli*, (B) mutant A, and (C) mutant B networks that maximize biomass yield. Results for mutants A and B assume the reactions responsible for 1,3-propanediol production are available.

Figure 5 shows 1,3-propanediol (PDO) production limits under aerobic conditions for mutant A (dash/dotted line), mutant B (dotted line), and the wild-type *E. coli* network (solid). The yellow points depict the solution identified by OptKnock (i.e., maximum chemical production at the maximum biomass yield).

Figure 6 shows projection of the multidimensional flux space onto two dimensions. The pink region represents flux ranges potentially reachable by both the mutant and complete networks, while the blue region corresponds to flux distributions rendered unreachable by the gene deletion(s). Point A represents the maximum biomass yield solution. Point B is the solution assuming the minimization of metabolic adjustment hypothesis for the mutant network, while point C is the solution assuming the mutant network will maximize its biomass yield.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The ability to investigate the metabolism of single-cellular organisms at a genomic scale, and thus systemic level, motivates the need for novel computational methods aimed at identifying strain engineering strategies. The present invention includes a computational framework termed OptKnock for suggesting gene deletion strategies leading to the overproduction of specific chemical compounds in *E. coli*. This is accomplished by ensuring that the production of the desired chemical becomes an obligatory byproduct of growth by "shaping" the connectivity of the metabolic network. In other words, OptKnock identifies and subsequently removes metabolic reactions that are capable of uncoupling cellular growth from chemical production. The computational procedure is designed to

identify not just straightforward but also non-intuitive knockout strategies by simultaneously considering the entire *E. coli* metabolic network as abstracted in the *in silico E. coli* model of Palsson and coworkers (Edwards and Palsson, 2000). The complexity and built-in redundancy of this network (e.g., the *E. coli* model encompasses 720 reactions) necessitates a systematic and efficient search approach to combat the combinatorial explosion of candidate gene knockout strategies.

The nested optimization framework shown in Figure 1 is developed to identify multiple gene deletion combinations that maximally couple cellular growth objectives with externally imposed chemical production targets. This multi-layered optimization structure involving two competing optimal strategists (i.e., cellular objective and chemical production) is referred to as a bilevel optimization problem (Bard, 1998). Problem formulation specifics along with an elegant solution procedure drawing upon linear programming (LP) duality theory are described in the Methods section. The OptKnock procedure is applied to succinate, lactate, and 1,3-propanediol (PDO) production in *E. coli* with the maximization of the biomass yield for a fixed amount of uptaken glucose employed as the cellular objective. The obtained results are also contrasted against using the minimization of metabolic adjustment (MOMA) (Segre et al., 2002) as the cellular objective. Based on the OptKnock framework, it is possible identify the most promising gene knockout strategies and their corresponding allowable envelopes of chemical versus biomass production in the context of succinate, lactate, and PDO production in *E. coli*.

A preferred embodiment of this invention describes a computational framework, termed OptKnock, for suggesting gene deletions strategies that could lead to chemical production in *E. coli* by ensuring that the drain towards metabolites/compounds necessary for growth resources (i.e., carbons, redox potential, and energy) must be accompanied, due to stoichiometry, by the production of the desired chemical. Therefore, the production of the desired product becomes an obligatory byproduct of cellular growth. Specifically, OptKnock pinpoints which reactions to remove from a metabolic network, which can be realized by deleting the gene(s) associated with the identified functionality. The procedure was demonstrated based on succinate, lactate, and PDO production in *E. coli* K-12. The obtained results exhibit good agreement with strains published in the literature. While some of the suggested gene deletions are quite straightforward, as they essentially prune

reaction pathways competing with the desired one, many others are at first quite non-intuitive reflecting the complexity and built-in redundancy of the metabolic network of *E. coli*. For the succinate case, OptKnock correctly suggested anaerobic fermentation and the removal of the phosphotranferase glucose uptake mechanism as a consequence of the competition between the cellular and chemical production objectives, and not as a direct input to the problem. In the lactate study, the glucokinase-based glucose uptake mechanism was shown to decouple lactate and biomass production for certain knockout strategies. For the PDO case, results show that the Entner-Doudoroff pathway is more advantageous than EMP glycolysis despite the fact that it is substantially less energetically efficient. In addition, the so far popular *tpi* knockout was clearly shown to reduce the maximum yields of PDO while a complex network of 15 reactions was shown to be theoretically possible of "leaking" flux from the PPP pathway to the TCA cycle and thus decoupling PDO production from biomass formation. The obtained results also appeared to be quite robust with respect to the choice for the cellular objective.

The present invention contemplates any number of cellular objectives, including but not limited to maximizing a growth rate, maximizing ATP production, minimizing metabolic adjustment, minimizing nutrient uptake, minimizing redox production, minimizing a Euclidean norm, and combinations of these and other cellular objectives.

It is important to note that the suggested gene deletion strategies must be interpreted carefully. For example, in many cases the deletion of a gene in one branch of a branched pathway is equivalent with the significant up-regulation in the other. In addition, inspection of the flux changes before and after the gene deletions provides insight as to which genes need to be up or down-regulated. Lastly, the problem of mapping the set of identified reactions targeted for removal to its corresponding gene counterpart is not always uniquely specified. Therefore, careful identification of the most economical gene set accounting for isozymes and multifunctional enzymes needs to be made.

Preferably, in the OptKnock framework, the substrate uptake flux (i.e., glucose) is assumed to be 10 mmol/gDW·hr. Therefore, all reported chemical production and biomass formation values are based upon this postulated and not predicted uptake scenario. Thus, it is quite possible that the suggested deletion mutants may involve substantially lower uptake efficiencies. However, because OptKnock essentially suggests mutants with coupled

growth and chemical production, one could envision a growth selection system that will successively evolve mutants with improved uptake efficiencies and thus enhanced desired chemical production characteristics.

Where there is a lack of any regulatory or kinetic information within the purely stoichiometric representation of the inner optimization problem that performs flux allocation, OptKnock is used to identify any gene deletions as the sole mechanism for chemical overproduction. Clearly, the lack of any regulatory or kinetic information in the model is a simplification that may in some cases suggest unrealistic flux distributions. The incorporation of regulatory information will not only enhance the quality of the suggested gene deletions by more appropriately resolving flux allocation, but also allow us to suggest regulatory modifications along with gene deletions as mechanisms for strain improvement. The use of alternate modeling approaches (e.g., cybernetic (Kompala et al., 1984; Ramakrishna et al., 1996; Varner and Ramkrishna, 1999), metabolic control analysis (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Hatzimanikatis et al., 1998)), if available, can be incorporated within the OptKnock framework to more accurately estimate the metabolic flux distributions of gene-deleted metabolic networks. Nevertheless, even without such regulatory or kinetic information, OptKnock provides useful suggestions for strain improvement and more importantly establishes a systematic framework. The present invention naturally contemplates future improvements in metabolic and regulatory modeling frameworks.

Methods

The maximization of a cellular objective quantified as an aggregate reaction flux for a steady state metabolic network comprising a set $\mathcal{N} = \{1,\ldots, N\}$ of metabolites and a set $\mathcal{M} = \{1,\ldots, M\}$ of metabolic reactions fueled by a glucose substrate is expressed mathematically as follows,

$$\text{maximize} \quad v_{\text{cellular objective}} \qquad \text{(Primal)}$$

$$\text{subject to} \quad \sum_{j=1}^{M} S_{ij} v_j = 0, \qquad \forall\, i \in \mathcal{N}$$

$$v_{pts} + v_{glk} = v_{glc\_uptake} \ mmol/gDW \cdot hr$$

$$v_{atp} \geq v_{atp\_main} \ mmol/gDW \cdot hr$$

$$v_{biomass} \geq v_{biomass}^{target} \ 1/hr$$

$$v_j \geq 0, \qquad \forall j \in \mathcal{M}_{irrev}$$

$$v_j \leq 0, \qquad \forall j \in \mathcal{M}_{secr\_only}$$

$$v_j \in \mathcal{R}, \qquad \forall j \in \mathcal{M}_{rev}$$

where S_{ij} is the stoichiometric coefficient of metabolite i in reaction j, v_j represents the flux of reaction j, v_{glc_uptake} is the basis glucose uptake scenario, v_{atp_main} is the non-growth associated ATP maintenance requirement, and $v_{biomass}^{target}$ is a minimum level of biomass production. The vector v includes both internal and transport reactions. The forward (i.e., positive) direction of transport fluxes corresponds to the uptake of a particular metabolite, whereas the reverse (i.e., negative) direction corresponds to metabolite secretion. The uptake of glucose through the phosphotransferase system and glucokinase are denoted by v_{pts} and v_{glk}, respectively. Transport fluxes for metabolites that can only be secreted from the network are members of $\mathcal{M}_{secr\_only}$. Note also that the complete set of reactions $\mathcal{M}$ is subdivided into reversible $\mathcal{M}_{rev}$ and irreversible $\mathcal{M}_{irrev}$ reactions. The cellular objective is often assumed to be a drain of biosynthetic precursors in the ratios required for biomass formation (Neidhardt and Curtiss, 1996). The fluxes are reported per 1 $gDW \cdot hr$ such that biomass formation is expressed as *g biomass produced/gDW·hr* or *1/hr*.

The modeling of gene deletions, and thus reaction elimination, first requires the incorporation of binary variables into the flux balance analysis framework (Burgard and Maranas, 2001; Burgard et al., 2001). These binary variables,

$$y_j = \begin{cases} 1 & \text{if reaction flux } v_j \text{ is active} \\ 0 & \text{if reaction flux } v_j \text{ is not active} \end{cases}, \forall j \in \mathcal{M}$$

assume a value of one if reaction j is active and a value of zero if it is inactive. The following constraint,

$$v_j^{min} \cdot y_j \leq v_j \leq v_j^{max} \cdot y_j \qquad , \forall j \in \mathcal{M}$$

ensures that reaction flux v_j is set to zero only if variable y_j is equal to zero. Alternatively, when y_j is equal to one, v_j is free to assume any value between a lower v_j^{min} and an upper v_j^{max} bound. In this study, v_j^{min} and v_j^{max} are identified by minimizing and subsequently maximizing every reaction flux subject to the constraints from the Primal problem.

The identification of optimal gene/reaction knockouts requires the solution of a bilevel optimization problem that chooses the set of reactions that can be accessed ($y_j = 1$) so as the optimization of the cellular objective indirectly leads to the overproduction of the chemical or biochemical of interest (see also Figure 1). Using biomass formation as the cellular objective, this is expressed mathematically as the following bilevel mixed-integer optimization problem

$$\underset{y_j}{\text{maximize}} \quad v_{chemical} \qquad \text{(OptKnock)}$$

$$\text{subject to} \quad \left(\begin{array}{lll} \underset{v_j}{\text{maximize}} & v_{biomass} & \text{(Primal)} \\ \text{subject to} & \sum_{j=1}^{M} S_{ij} v_j = 0, & \forall i \in \mathcal{N} \\ & v_{pts} + v_{glk} = v_{glc\_uptake} & \\ & v_{atp} \geq v_{atp\_main} & \\ & v_{biomass} \geq v_{biomass}^{target} & \end{array} \right)$$

$$v_j^{min} \cdot y_j \leq v_j \leq v_j^{max} \cdot y_j, \qquad \forall j \in \mathcal{M}$$

$$y_j = \{0,1\}, \qquad \forall j \in \mathcal{M}$$

$$\sum_{j \in M} (1 - y_j) \leq K$$

where K is the number of allowable knockouts. The final constraint ensures that the resulting network meets a minimum biomass yield, $v_{biomass}^{target}$.

The direct solution of this two-stage optimization problem is intractable given the high dimensionality of the flux space (i.e., over 700 reactions) and the presence of two

nested optimization problems. To remedy this, we develop an efficient solution approach borrowing from LP duality theory which shows that for every linear programming problem (primal) there exists a unique optimization problem (dual) whose optimal objective value is equal to that of the primal problem. A similar strategy was employed by (Burgard and Maranas, 2003) for identifying/testing metabolic objective functions from metabolic flux data. The dual problem (Ignizio and Cavalier, 1994) associated with the OptKnock inner problem is

$$\text{minimize} \quad v_{atp\_main} \cdot \mu_{atp} + v_{biomass}^{target} \cdot \mu_{biomass} + v_{glc\_uptake} \cdot glc \qquad \text{(Dual)}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \lambda_i^{stoich} S_{i,glk} + \mu_{glk} + glc = 0$$

$$\sum_{i=1}^{N} \lambda_i^{stoich} S_{i,pts} + \mu_{pts} + glc = 0$$

$$\sum_{i=1}^{N} \lambda_i^{stoich} S_{i,biomass} + \mu_{biomass} = 1$$

$$\sum_{i=1}^{N} \lambda_i^{stoich} S_{ij} + \mu_j = 0, \qquad \forall j \in \mathcal{M},\ j \neq glk, pts, biomass$$

$$\mu_j^{min} \cdot (1 - y_j) \leq \mu_j \leq \mu_j^{max} \cdot (1 - y_j), \quad \forall j \in \mathcal{M}_{irrev} \text{ and } j \notin \mathcal{M}_{secr\_only}$$

$$\mu_j \geq \mu_j^{min} \cdot (1 - y_j), \qquad \forall j \in \mathcal{M}_{irrev} \text{ and } \mathcal{M}_{secr\_only}$$

$$\mu_j \leq \mu_j^{max} \cdot (1 - y_j), \qquad \forall j \in \mathcal{M}_{irrev} \text{ and } j \notin \mathcal{M}_{secr\_only}$$

$$\mu_j \in \mathcal{R}, \qquad \forall j \in \mathcal{M}_{irrev} \text{ and } \mathcal{M}_{secr\_only}$$

$$\lambda_i^{stoich} \in \mathcal{R}, \qquad \forall j \in \mathcal{N}$$

$$glc \in \mathcal{R}$$

where λ_i^{stoich} is the dual variable associated with the stoichiometric constraints, glc is the dual variable associated with the glucose uptake constraint, and μ_j is the dual variable associated with any other restrictions on its corresponding flux v_j in the Primal. Note that the dual variable μ_j acquires unrestricted sign if its corresponding flux in the OptKnock inner problem is set to zero by enforcing $y_j = 0$. The parameters μ_j^{min} and μ_j^{max} are identified

by minimizing and subsequently maximizing their values subject to the constraints of the Dual problem.

If the optimal solutions to the Primal and Dual problems are bounded, their objective function values must be equal to one another at optimality. This means that every optimal solution to both problems can be characterized by setting their objectives equal to one another and accumulating their respective constraints. Thus the bilevel formulation for OptKnock shown previously can be transformed into the following single-level MILP

maximize $v_{chemical}$ (OptKnock)

subject to

$$v_{biomass} = v_{atp\_main} \cdot \mu_{atp} + v_{biomass}^{target} \cdot \mu_{biomass} + v_{glc\_uptake} \cdot glc$$

$$\sum_{j=1}^{M} S_{ij} v_j = 0, \qquad \forall i \in \mathcal{N}$$

$$v_{pts} + v_{glk} = v_{glc\_uptake} \ mmol/gDW\,hr$$

$$v_{atp} \geq v_{atp\_main} \ mmol/gDW\,hr$$

$$\sum_{i=1}^{N} \lambda_i^{stoich} S_{i,glk} + \mu_{glk} + glc = 0$$

$$\sum_{i=1}^{N} \lambda_i^{stoich} S_{i,pts} + \mu_{pts} + glc = 0$$

$$\sum_{i=1}^{N} \lambda_i^{stoich} S_{i,biomass} + \mu_{biomass} = 1$$

$$\sum_{i=1}^{N} \lambda_i^{stoich} S_{ij} + \mu_j = 0, \qquad \forall j \in \mathcal{M},\ j \neq glk, pts, biomass$$

$$\sum_{j \in M} (1 - y_j) \leq K$$

$$v_{biomass} \geq v_{biomass}^{target}$$

$$\mu_j^{\min} \cdot (1 - y_j) \le \mu_j \le \mu_j^{\max} \cdot (1 - y_j), \quad \forall j \in \mathcal{M}_{\text{rev}} \text{ and } j \notin \mathcal{M}_{\text{secr\_only}}$$

$$\mu_j \ge \mu_j^{\min} \cdot (1 - y_j), \quad \forall j \in \mathcal{M}_{\text{rev}} \text{ and } \mathcal{M}_{\text{secr\_only}}$$

$$\mu_j \le \mu_j^{\max} \cdot (1 - y_j), \quad \forall j \in \mathcal{M}_{\text{irrev}} \text{ and } j \notin \mathcal{M}_{\text{secr\_only}}$$

$$\mu_j \in \mathcal{R}_i \quad \forall j \in \mathcal{M}_{\text{irrev}} \text{ and } \mathcal{M}_{\text{secr\_only}}$$

$$v_j^{\min} \cdot y_j \le v_j \le v_j^{\max} \cdot y_j, \quad \forall j \in \mathcal{M}$$

$$\lambda_i^{\text{stoich}} \in \mathcal{R}_i \quad \forall j \in \mathcal{N}$$

$$\text{glc} \in \mathcal{R}$$

$$y_j = \{0,1\}, \quad \forall j \in \mathcal{M}$$

An important feature of the above formulation is that if the problem is feasible, the optimal solution will always be found. In this invention, the candidates for gene knockouts included, but are not limited to, all reactions of glycolysis, the TCA cycle, the pentose phosphate pathway, respiration, and all anaplerotic reactions. This is accomplished by limiting the number of reactions included in the summation (i.e., $\sum_{j \in \text{Central Metabolism}} (1 - y_j) = K$).

Problems containing as many as 100 binary variables were solved in the order of minutes to hours using CPLEX 7.0 accessed via the GAMS modeling environment on an IBM RS6000-270 workstation. It should be understood, however, that the present invention is not dependent upon any particular type of computer or environment being used. Any type can be used to allow for inputting and outputting the information associated with the methodology of the present invention. Moreover, the steps of the methods of the present invention can be implemented in any number of types software applications, or languages, and the present invention is not limited in this respect.

It will be appreciated that other embodiments and uses will be apparent to those skilled in the art and that the invention is not limited to these specific illustrative examples.

EXAMPLE I

Succinate and Lactate Production

Which reactions, if any, that could be removed from the *E. coli* K-12 stoichiometric model (Edwards and Palsson, 2000) so as the remaining network produces succinate or lactate whenever biomass maximization is a good descriptor of flux allocation were identified. A prespecified amount of glucose (10 mmol/gDW·hr), along with unconstrained uptake routes for inorganic phosphate, oxygen, sulfate, and ammonia are provided to fuel the metabolic network. The optimization step could opt for or against the phosphotransferase system, glucokinase, or both mechanisms for the uptake of glucose. Secretion routes for acetate, carbon dioxide, ethanol, formate, lactate and succinate are also enabled. Note that because the glucose uptake rate is fixed, the biomass and product yields are essentially equivalent to the rates of biomass and product production, respectively. In all cases, the OptKnock procedure eliminated the oxygen uptake reaction pointing at anaerobic growth conditions consistent with current succinate (Zeikus et al., 1999) and lactate (Datta et al., 1995) fermentative production strategies.

Table I summarizes three of the identified gene knockout strategies for succinate overproduction (*i.e.*, mutants A, B, and C). The anaerobic flux distributions at the maximum biomass yields for the complete *E. coli* network (i.e., wild-type), mutant B, and mutant C are illustrated in Figure 2A-C. The results for mutant A suggested that the removal of two reactions (i.e., pyruvate formate lyase and lactate dehydrogenase) from the network results in succinate production reaching 63% of its theoretical maximum at the maximum biomass yield. This knockout strategy is identical to the one employed by Stols and Donnelly (Stols and Donnelly, 1997) in their succinate overproducing *E. coli* strain. Next, the envelope of allowable succinate versus biomass production was explored for the wild-type *E. coli* network and the three mutants listed in Table I. Note that the succinate production limits, shown in Figure 3A, revealed that mutant A does not exhibit coupled succinate and biomass formation until the yield of biomass approaches 80% of the maximum. Mutant B, however, with the additional deletion of acetaldehyde dehydrogenase, resulted in a much earlier coupling of succinate with biomass yields.

A less intuitive strategy was identified for mutant C which focused on inactivating two PEP consuming reactions rather than eliminating competing byproduct (i.e., ethanol,

formate, and lactate) production mechanisms. First, the phosphotransferase system was disabled requiring the network to rely exclusively on glucokinase for the uptake of glucose. Next, pyruvate kinase was removed leaving PEP carboxykinase as the only central metabolic reaction capable of draining the significant amount of PEP supplied by glycolysis. This strategy, assuming that the maximum biomass yield could be attained, resulted in a succinate yield approaching 88% of the theoretical maximum. In addition, Figure 3A revealed significant succinate production for every attainable biomass yield, while the maximum theoretical yield of succinate is the same as that for the wild-type strain.

The OptKnock framework was next applied to identify knockout strategies for coupling lactate and biomass production. Table I shows three of the identified gene knockout strategies (*i.e.*, mutants A, B, and C) and the flux distribution of mutant C at the maximum biomass yield is shown in Figure 2D. Mutant A redirects flux toward lactate at the maximum biomass yield by blocking acetate and ethanol production. This result is consistent with previous work demonstrating that an *adh, pta* mutant *E. coli* strain could grow anaerobically on glucose by producing lactate (Gupta and Clark, 1989). Mutant B provides an alternate strategy involving the removal of an initial glycolysis reaction along with the acetate production mechanism. This results in a lactate yield of 90% of its theoretical limit at the maximum biomass yield. The vertical red line for mutant B in Figure 3B indicates that the network could avoid producing lactate while maximizing biomass formation. This is due to the fact that OptKnock does not explicitly account for the "worst-case" alternate solution. It should be appreciated that upon the additional elimination of the glucokinase and ethanol production mechanisms, mutant C exhibited a tighter coupling between lactate and biomass production.

Table I – Biomass and chemical yields for various gene knockout strategies identified by OptKnock. The reactions and corresponding enzymes for each knockout strategy are listed. The maximum biomass and corresponding chemical yields are provided on a basis of 10 mmol/hr glucose fed and 1 gDW of cells. The rightmost column provides the chemical yields for the same basis assuming a minimal redistribution of metabolic fluxes from the wild-type (undeleted) *E. coli* network (MOMA assumption). For the 1,3-propanediol case, glycerol secretion was disabled for both knockout strategies.

Succinate				$max\ v_{biomass}$		$min\ \Sigma(v_o - v)^2$
				Biomass	Succinate	Succinate
ID		Knockouts	Enzyme	(1/hr)	(mmol/hr)	(mmol/hr)
Wild		"Complete network"		0.38	0.12	0
A	1	COA + PYR → ACCOA + FOR	Pyruvate formate lyase	0.31	10.70	1.65
	2	NADH + PYR ↔ LAC + NAD	Lactate dehydrogenase			
B	1	COA + PYR → ACCOA + FOR	Pyruvate formate lyase			
	2	NADH + PYR ↔ LAC + NAD	Lactate dehydrogenase	0.31	10.70	4.79
	3	ACCOA + 2 NADH ↔ COA + ETH + 2 NAD	Acetaldehyde dehydrogenase			
C	1	ADP + PEP → ATP + PYR	Pyruvate kinase			
	2	ACTP + ADP ↔ AC + ATP or	Acetate kinase	0.16	15.15	6.21
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase			
	3	GLC + PEP → G6P + PYR	Phosphotransferase system			

Lactate				$max\ v_{biomass}$		$min\ \Sigma(v_o - v)^2$
				Biomass	Lactate	Lactate
ID		Knockouts	Enzyme	(1/hr)	(mmol/hr)	(mmol/hr)
Wild		"Complete network"		0.38	0	0
A	1	ACTP + ADP ↔ AC + ATP or	Acetate kinase			
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase	0.28	10.46	5.58
	2	ACCOA + 2 NADH ↔ COA + ETH + 2 NAD	Acetaldehyde dehydrogenase			
B	1	ACTP + ADP ↔ AC + ATP or	Acetate kinase			
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase	0.13	18.00	0.19
	2	ATP + F6P → ADP + FDP or	Phosphofructokinase			
		FDP ↔ T3P1 + T3P2	Fructose-1,6-bisphosphatate aldolase			
C	1	ACTP + ADP ↔ AC + ATP or	Acetate kinase			
		ACCOA + Pi ↔ ACTP + COA	Phosphotransacetylase			
	2	ATP + F6P → ADP + FDP or	Phosphofructokinase	0.12	18.13	10.53
		FDP ↔ T3P1 + T3P2	Fructose-1,6-bisphosphatate aldolase			
	3	ACCOA + 2 NADH ↔ COA + ETH + 2 NAD	Acetaldehyde dehydrogenase			
	4	GLC + ATP → G6P + PEP	Glucokinase			

1,3-Propanediol				$max\ v_{biomass}$		$min\ \Sigma(v_o - v)^2$
				Biomass	1,3-PD	1,3-PD
ID		Knockouts	Enzyme	(1/hr)	(mmol/hr)	(mmol/hr)
Wild		"Complete network"		1.06	0	0
A	1	FDP → F6P + Pi or	Fructose-1,6-bisphosphatase			
		FDP ↔ T3P1 + T3P2	Fructose-1,6-bisphosphate aldolase			
	2	13PDG + ADP ↔ 3PG + ATP or	Phosphoglycerate kinase	0.21	9.66	8.66
		NAD + Pi + T3P1 ↔ 13PDG + NADH	Glyceraldehyde-3-phosphate dehydrogenase			
	3	GL + NAD ↔ GLAL + NADH	Aldehyde dehydrogenase			
B	1	T3P1 ↔ T3P2	Triosphosphate isomerase			
	2	G6P + NADP ↔ D6PGL + NADPH or	Glucose 6-phosphate-1-dehydrogenase			
		D6PGL → D6PGC	6-Phosphogluconolactonase	0.29	9.67	9.54
	3	DRSP → ACAL + T3P1	Deoxyribose-phosphate aldolase			
	4	GL + NAD ↔ GLAL + NADH	Aldehyde dehydrogenase			

EXAMPLE 2

1,3-Propanediol (PDO) Production

In addition to devising optimum gene knockout strategies, OptKnock was used to design strains where gene additions were needed along with gene deletions such as in PDO production in *E. coli*. Although microbial 1,3-propanediol (PDO) production methods have been developed utilizing glycerol as the primary carbon source (Hartlep et al., 2002; Zhu et al., 2002), the production of 1,3-propanediol directly from glucose in a single microorganism has recently attracted considerable interest (Cameron et al., 1998; Biebl et al., 1999; Zeng and Biebl, 2002). Because wild-type *E. coli* lacks the pathway necessary for PDO production, the gene addition framework was first employed (Burgard and Maranas, 2001) to identify the additional reactions needed for producing PDO from glucose in *E. coli*. The gene addition framework identified a straightforward three-reaction pathway involving the conversion of glycerol-3-P to glycerol by glycerol phosphatase, followed by the conversion of glycerol to 1,3 propanediol by glycerol dehydratase and 1,3-propanediol oxidoreductase. These reactions were then added to the *E. coli* stoichiometric model and the OptKnock procedure was subsequently applied.

OptKnock revealed that there was neither a single nor a double deletion mutant with coupled PDO and biomass production. However, one triple and multiple quadruple knockout strategies that can couple PDO production with biomass production was identified. Two of these knockout strategies are shown in Table I. The results suggested that the removal of certain key functionalities from the *E. coli* network resulted in PDO overproducing mutants for growth on glucose. Specifically, Table I reveals that the removal of two glycolytic reactions along with an additional knockout preventing the degradation of glycerol yields a network capable of reaching 72% of the theoretical maximum yield of PDO at the maximum biomass yield. Note that the glyceraldehyde-3-phosphate dehydrogenase (*gapA*) knockout was used by DuPont in their PDO-overproducing *E. coli* strain (Nakamura, 2002). Mutant B revealed an alternative strategy, involving the removal of the triose phosphate isomerase (*tpi*) enzyme exhibiting a similar PDO yield and a 38% higher biomass yield. Interestingly, a yeast strain deficient in triose phosphate isomerase activity was recently reported to produce glycerol, a key precursor to PDO, at 80-90% of its maximum theoretical yield (Compagno et al., 1996).

The flux distributions of the wild-type *E. coli*, mutant A, and mutant B networks that maximize the biomass yield are available in Figure 4. Not surprisingly, further conversion of glycerol to glyceraldehyde was disrupted in both mutants A and B. For mutant A, the removal of two reactions from the top and bottom parts of glycolysis resulted in a nearly complete inactivation of the pentose phosphate and glycolysis (with the exception of triose phosphate isomerase) pathways. To compensate, the Entner-Doudoroff glycolysis pathway is activated to channel flux from glucose to pyruvate and glyceraldehyde-3-phosphate (GAP). GAP is then converted to glycerol which is subsequently converted to PDO. Energetic demands lost with the decrease in glycolytic fluxes from the wild-type *E. coli* network case, are now met by an increase in the TCA cycle fluxes. The knockouts suggested for mutant B redirect flux toward the production of PDO by a distinctly different mechanism. The removal of the initial pentose phosphate pathway reaction results in the complete flow of metabolic flux through the first steps of glycolysis. At the fructose bisphosphate aldolase junction, the flow is split into the two product metabolites: dihydroxyacetone-phosphate (DHAP) which is converted to PDO and GAP which continues through the second half of the glycolysis. The removal of the triose-phosphate isomerase reaction prevents any interconversion between DHAP and GAP. Interestingly, a fourth knockout is predicted to retain the coupling between biomass formation and chemical production. This knockout prevents the "leaking" of flux through a complex pathway involving 15 reactions that together convert ribose-5-phosphate (R5P) to acetate and GAP, thereby decoupling growth from chemical production.

Next, the envelope of allowable PDO production versus biomass yield is explored for the two mutants listed in Table I. The production limits of the mutants along with the original *E. coli* network, illustrated in Figure 5, reveal that the wild-type *E. coli* network has no "incentive" to produce PDO if the biomass yield is to be maximized. On the other hand, both mutants A and B have to produce significant amounts of PDO if any amount of biomass is to be formed given the reduced functionalities of the network following the gene removals. Mutant A, by avoiding the *tpi* knockout that essentially sets the ratio of biomass to PDO production, is characterized by a higher maximum theoretical yield of PDO. The above described results hinge on the use of glycerol as a key intermediate to PDO. Next,

the possibility of utilizing an alternative to the glycerol conversion route for 1,3-propandediol production was explored.

Applicants identified a pathway in *Chloroflexus aurantiacus* involving a two-step NADPH-dependant reduction of malonyl-CoA to generate 3-hydroxypropionic acid (3-HPA) (Menendez et al., 1999; Hugler et al., 2002). 3-HPA could then be subsequently converted chemically to 1,3 propanediol given that there is no biological functionality to achieve this transformation. This pathway offers a key advantage over PDO production through the glycerol route because its initial step (acetyl-CoA carboxylase) is a carbon fixing reaction. Accordingly, the maximum theoretical yield of 3-HPA (1.79 mmol/mmol glucose) is considerably higher than for PDO production through the glycerol conversion route (1.34 mmol/mmol glucose). The application of the OptKnock framework upon the addition of the 3-HPA production pathway revealed that many more knockouts are required before biomass formation is coupled with 3-HPA production. One of the most interesting strategies involves nine knockouts yielding 3-HPA production at 91% of its theoretical maximum at optimal growth. The first three knockouts were relatively straightforward as they involved removal of competing acetate, lactate, and ethanol production mechanisms. In addition, the Entner-Doudoroff pathway (either phosphogluconate dehydratase or 2-keto-3-deoxy-6-phosphogluconate aldolase), four respiration reactions (i.e., NADH dehydrogenase I, NADH dehydrogenase II, glycerol-3-phosphate dehydrogenase, and the succinate dehydrogenase complex), and an initial glycolyis step (i.e., phosphoglucose isomerase) are disrupted. This strategy resulted in a 3-HPA yield that, assuming the maximum biomass yield, is 69% higher than the previously identified mutants utilizing the glycerol conversion route.

EXAMPLE 3

Alternative Cellular Objective: Minimization of Metabolic Adjustment

All results described previously were obtained by invoking the maximization of biomass yield as the cellular objective that drives flux allocation. This hypothesis essentially assumes that the metabolic network could arbitrarily change and/or even rewire regulatory loops to maintain biomass yield maximality under changing environmental conditions (maximal response). Recent evidence suggests that this is sometimes achieved by the K-12 strain of *E. coli* after multiple cycles of growth selection (Ibarra et al., 2002).

In this section, a contrasting hypothesis was examined (i.e., minimization of metabolic adjustment (MOMA) (Segre et al., 2002)) that assumed a myopic (minimal) response by the metabolic network upon gene deletions. Specifically, the MOMA hypothesis suggests that the metabolic network will attempt to remain as close as possible to the original steady state of the system rendered unreachable by the gene deletion(s). This hypothesis has been shown to provide a more accurate description of flux allocation immediately after a gene deletion event (Segre et al., 2002). Figure 6 pictorially shows the two differing new steady states predicted by the two hypotheses, respectively. For this study, the MOMA objective was utilized to predict the flux distributions in the mutant strains identified by OptKnock. The base case for the lactate and succinate simulations was assumed to be maximum biomass formation under anaerobic conditions, while the base case for the PDO simulations was maximum biomass formation under aerobic conditions. The results are shown in the last column of Table 1. In all cases, the suggested multiple gene knock-out strategy suggests only slightly lower chemical production yields for the MOMA case compared to the maximum biomass hypothesis. This implies that the OptKnock results are fairly robust with respect to the choice of cellular objective.

The publications and other material used herein to illuminate the background of the invention or provide additional details respecting the practice, are herein incorporated by reference in their entirety.

The present invention contemplates numerous variations, including variations in organisms, variations in cellular objectives, variations in bioengineering objectives, variations in types of optimization problems formed and solutions used.

These and/or other variations, modifications or alterations may be made therein without departing from the spirit and the scope of the invention as set forth in the appended claims.

REFERENCES

Badarinarayana, V, Estep, PW, 3rd, Shendure, J, Edwards, J, Tavazoie, S, Lam, F, Church, GM. 2001. Selection analyses of insertional mutants using subgenic-resolution arrays. Nat Biotechnol 19(11): 1060-5.

Bard, JF. 1998. Practical bilevel optimization : algorithms and applications. Dordrecht ; Boston, Kluwer Academic.

Biebl, H, Menzel, K, Zeng, AP, Deckwer, WD. 1999. Microbial production of 1,3-propanediol. Appl Environ Microbiol 52: 289-297.

Burgard, AP, Maranas, CD. 2001. Probing the performance limits of the Escherichia coli metabolic network subject to gene additions or deletions. Biotechnol Bioeng 74: 364-375.

Burgard, AP, Maranas, CD. 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng 82(6): 670-7.

Burgard, AP, Vaidyaraman, S, Maranas, CD. 2001. Minimal Reaction Sets for *Escherichia coli* Metabolism under Different Growth Requirements and Uptake Environments. Biotechnol Prog 17: 791-797.

Cameron, DC, Altaras, NE, Hoffman, ML, Shaw, AJ. 1998. Metabolic engineering of propanediol pathways. Biotechnol Prog 14(1): 116-25.

Compagno, C, Boschi, F, Ranzi, BM. 1996. Glycerol production in a triose phosphate isomerase deficient mutant of Saccharomyces cerevisiae. Biotechnol Prog 12(5): 591-5.

Covert, MW, Palsson, BO. 2002. Transcriptional regulation in constraints-based metabolic models of Escherichia coli. J Biol Chem 277(31): 28058-64.

Covert, MW, Schilling, CH, Palsson, BO. 2001. Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213(1): 73-88.

Datta, R, Tsai, S, Bonsignore, P, Moon, S, Frank, JR. 1995. Technological and economic potential of poly(lactic acid) and lactic acid derivatives. FEMS Microbiol. Rev. 16: 221-231.

Edwards, JS, Ibarra, RU, Palsson, BO. 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19(2): 125-30.

Edwards, JS, Palsson, BO. 2000. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 97(10): 5528-33.

Forster, J, Famili, I, Fu, PC, Palsson, B, Nielsen, J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Research 13(2): 244-253.

Gupta, S, Clark, DP. 1989. Escherichia coli derivatives lacking both alcohol dehydrogenase and phosphotransacetylase grow anaerobically by lactate fermentation. J Bacteriol 171(7): 3650-5.

Hartlep, M, Hussmann, W, Prayitno, N, Meynial-Salles, I, Zeng, AP. 2002. Study of two-stage processes for the microbial production of 1,3-propanediol from glucose. Appl Microbiol Biotechnol 60(1-2): 60-6.

Hatzimanikatis, V, Emmerling, M, Sauer, U, Bailey, JE. 1998. Application of mathematical tools for metabolic design of microbial ethanol production. Biotechnol Bioeng 58(2-3): 154-61.

Heinrich, R, Rapoport, TA. 1974. A linear steady-state treatment of enzymatic chains. Eur. J. Biochem. 41: 89-95.

Hugler, M, Menendez, C, Schagger, H, Fuchs, G. 2002. Malonyl-coenzyme A reductase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO(2) fixation. J Bacteriol 184(9): 2404-10.

Ibarra, RU, Edwards, JS, Palsson, BO. 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420(6912): 186-9.

Ignizio, JP, Cavalier, TM. 1994. Linear programming. Englewood Cliffs, N.J., Prentice Hall.

Kacser, H, Burns, JA. 1973. The control of flux. Symp. Soc. Exp. Biol. 27: 65-104.

Kompala, DS, Ramkrishna, D, Tsao, GT. 1984. Cybernetic Modeling of Microbial Growth on Multiple Substrates. Biotechnol Bioeng 26(11): 1272-1281.

Majewski, RA, Domach, MM. 1990. Simple constrained optimization view of acetate overflow in *Escherichia coli*. Biotechnol Bioeng 35: 732-738.

Menendez, C, Bauer, Z, Huber, H, Gad'on, N, Stetter, KO, Fuchs, G. 1999. Presence of acetyl coenzyme A (CoA) carboxylase and propionyl-CoA carboxylase in autotrophic Crenarchaeota and indication for operation of a 3-hydroxypropionate cycle in autotrophic carbon fixation. J Bacteriol 181(4): 1088-98.

Nakamura, CE. 2002. Production of 1,3-Propanediol by *E. coli*. presented at Metab Eng IV Conf: Tuscany, Italy.

Neidhardt, FC, Curtiss, R. 1996. *Escherichia coli* and *Salmonella* : cellular and molecular biology. Washington, D.C., ASM Press.

Papin, JA, Price, ND, Wiback, SJ, Fell, DA, Palsson, B. 2003. Metabolic Pathways in the Post-Genome Era. Trends Biochem Sci, accepted.

Price, ND, Papin, JA, Schilling, CH, Palsson, B. 2003. Genome-scale Microbial In Silico Models: The Constraints-Based Approach. Trends Biotechnol, accepted.

Ramakrishna, R, Edwards, JS, McCulloch, A, Palsson, BO. 2001. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. Am J Physiol Regul Integr Comp Physiol 280(3): R695-704.

Ramakrishna, R, Ramakrishna, D, Konopka, AE. 1996. Cybernetic Modeling of Growth in Mixed, Substitutable Substrate Environments: Preferential and Simultaneous Utilization. Biotechnol Bioeng 52: 141-151.

Schilling, CH, Covert, MW, Famili, I, Church, GM, Edwards, JS, Palsson, BO. 2002. Genome-scale metabolic model of Helicobacter pylori 26695. J Bacteriol 184(16): 4582-93.

Schilling, CH, Palsson, BO. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. J Theor Biol 203(3): 249-83.

Segre, D, Vitkup, D, Church, GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99(23): 15112-7.

Stephanopoulos, G, Aristidou, AA, Nielsen, J. 1998. Metabolic engineering : principles and methodologies. San Diego, Academic Press.

Stols, L, Donnelly, MI. 1997. Production of succinic acid through overexpression of NAD(+)-dependent malic enzyme in an Escherichia coli mutant. Appl Environ Microbiol 63(7): 2695-701.

Varma, A, Boesch, BW, Palsson, BO. 1993. Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. Appl Environ Microbiol 59(8): 2465-73.

Varma, A, Palsson, BO. 1993. Metabolic Capabilities of Escherichia coli: II. Optimal Growth Patterns. J. Theor. Biol. 165: 503-522.

Varma, A, Palsson, BO. 1994. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Bio/Technology 12: 994-998.

Varner, J, Ramkrishna, D. 1999. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. Biotechnol Prog 15(3): 407-25.

Zeikus, JG, Jain, MK, Elankovan, P. 1999. Biotechnology of succinate acid production and markets for derived industrial products. Appl Microbiol Biotechnol 51: 545-552.

Zeng, AP, Biebl, H. 2002. Bulk chemicals from biotechnology: the case of 1,3-propanediol production and the new trends. Adv Biochem Eng Biotechnol 74: 239-59.

Zhu, MM, Lawman, PD, Cameron, DC. 2002. Improving 1,3-propanediol production from glycerol in a metabolically engineered Escherichia coli by reducing accumulation of sn-glycerol-3-phosphate. Biotechnol Prog 18(4): 694-9.

What is claimed is:

1. A method for determining candidates for gene deletions and additions using a model of a metabolic network associated with an organism, the model comprising a plurality of metabolic reactions defining metabolite relationships, the method comprising: selecting a bioengineering objective for the organism; selecting at least one cellular objective; forming an optimization problem that couples the at least one cellular objective with the bioengineering objective; and solving the optimization problem to yield at least one candidate.

2. The method of claim 1 further comprising modifying the organism with the candidate.

3. The method of claim 1 wherein the bioengineering objective is overproduction of a chemical.

4. The method of claim 1 wherein the bioengineering objective is underproduction of a chemical.

5. The method of claim 1 wherein the cellular objective is growth.

6. The method of claim 1 wherein the cellular objective is minimization of metabolic adjustment.

7. The method of claim 1 wherein the candidate is a candidate for gene deletion, and the optimization problem includes a binary value for specifying if reaction flux is active or inactive.

8. The method of claim 1 wherein the optimization problem is a bilevel optimization problem.

9. The method of claim 1 wherein the optimization problem is a mixed-integer optimization problem.

10. The method of claim 1 wherein the optimization problem includes at least one stoichiometric constraint.

11. The method of claim 1 wherein the optimization problem includes at least one chemical uptake constraint.

12. The method of claim 1 wherein the step of forming an optimization problem includes quantifying the cellular objective as an aggregate reaction flux.

13. The method of claim 1 further comprising evaluating performance limits of the metabolic network with the at least one candidate based on ability of the network to meet the at least one cellular objective.

14. The method of claim 1 wherein the cellular objective is selected from the group consisting of: maximizing a growth rate, maximizing ATP production, minimizing metabolic adjustment, minimizing nutrient uptake, minimizing redox production, minimizing a Euclidean norm, and combinations thereof.

15. The method of claim 1 wherein the bioengineering objective is overproduction of glycerol and at least one candidate is for gene deletion and comprising genes coding for the enzymes fructose-1,6-bisphosphatase, fructose-1,6-bisphosphatase aldolase, phosphoglycerate kinase, glyceraldehydes-3-phosphate dehydrogenase, phosphoenolpyruvate synthase, NADH dehydrogenase I, phosphogluconate dehydratase, 2-keto-3-deoxy-6-phosphofluconate aldolase, triosphosphate isomerase, glucose 6-phosphate-1-dehydrogenase, 6-phosphogluconolactonase, deoxyribose-phosphate aldolase, aldehyde dehydrogenase, or combinations thereof.

16. The method of claim 1 wherein the bioengineering objective is overproduction of 1,3-propanediol and at least one candidate is for gene deletion and comprising genes coding for the enzymes fructose-1,6-bisphosphatase, fructose-1,6-bisphosphatase aldolase, phosphoglycerate kinase, glyceraldehyde-3-phosphate dehyrogenase, triosphosphate isomerase, glucose 6-phosphate-1-dehydrogenase, 6-phosphogluconolactonase, deoxyribose-phosphate aldolase, aldehyde dehydrogenase, or combinations thereof.

17. The method of claim 1 wherein the bioengineering objective is overproduction of succinate and at least one candidate is for gene deletion and comprising genes coding for the enzymes pyruvate formate lyase, acetaldehyde dehydrogenase, pyruvate kinase, FOF1-ATPase, NADH dehydrogenase I, fumarase, D-Lactate dehydrogenase, pyridine nucleotide transhydrogenase, phosphotransacetylase, acetate kinase, phosphotransferase, or combinations thereof.

18. The method of claim 1 wherein the bioengineering objective is overproduction of lactate and at least one candidate is for gene deletion and comprising genes coding for the enzymes phosphotransacetylase, acetate kinase, phosphofructokinase, fructose-1,6-bisphosphatase aldolase, triosphosphate isomerase, acetaldehyde dehyrogenase, glucokinase, or combinations thereof.

19. A computer-based method for determining candidates for gene deletions and additions using a model of a metabolic network associated with an organism, the model comprising a plurality of metabolic reactions defining metabolite relationships, the method comprising: inputting at least one bioengineering objective; receiving as input as least one cellular objective; forming an optimization that quantifies the cellular objective as an aggregate reaction flux and couples the at least one cellular objective with the bioengineering objective; solving the optimization problem to yield at least one candidate; and outputting the at least one candidate.

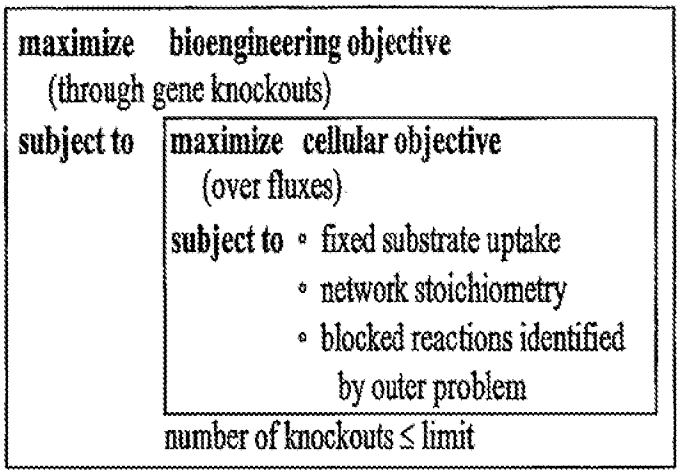
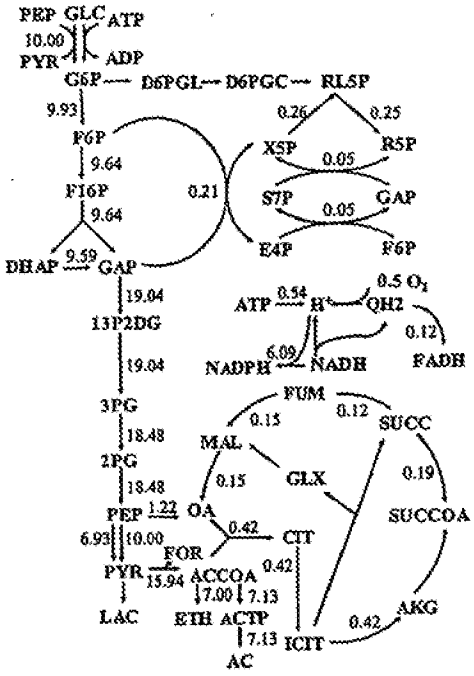
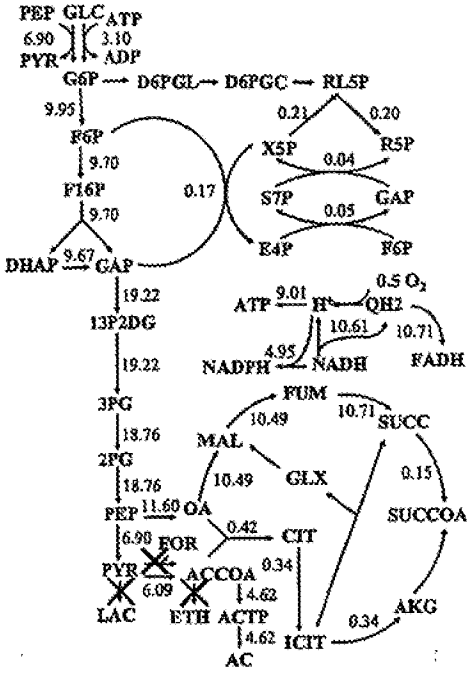

maximize bioengineering objective
(through gene knockouts)
subject to
maximize cellular objective
(over fluxes)
subject to
fixed substrate uptake
network stoichiometry
blocked reactions identified
by outer problem
number of knockouts ≤ limit

Figure 1

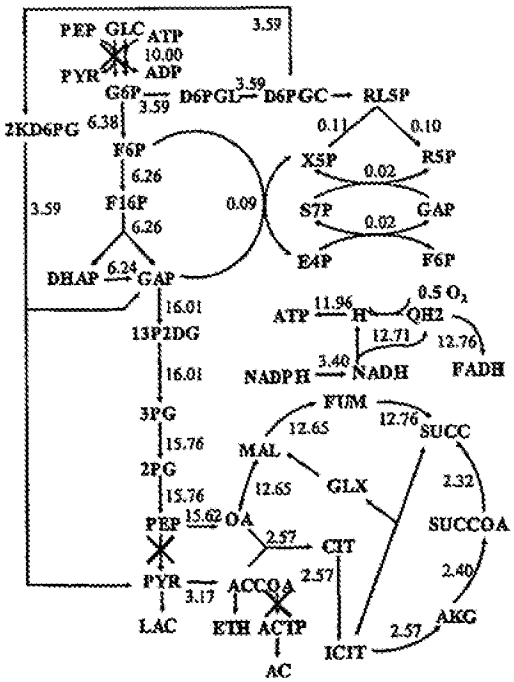
(A) "Wild-type" *E. coli*



(B) Succinate Mutant B



(C) Succinate Mutant C



(D) Lactate Mutant C

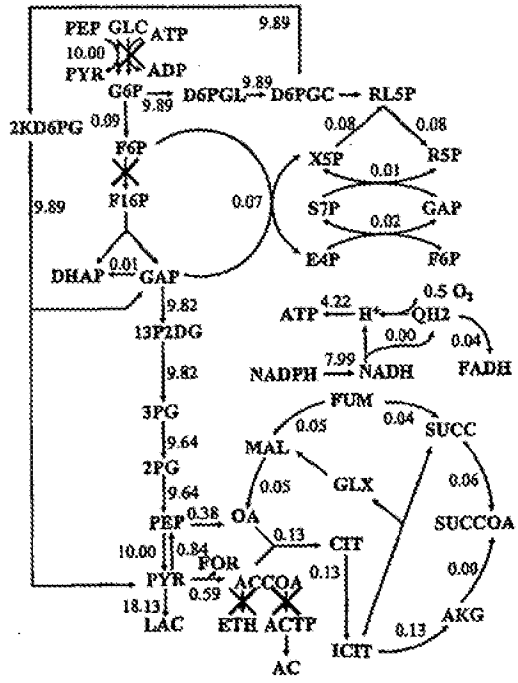
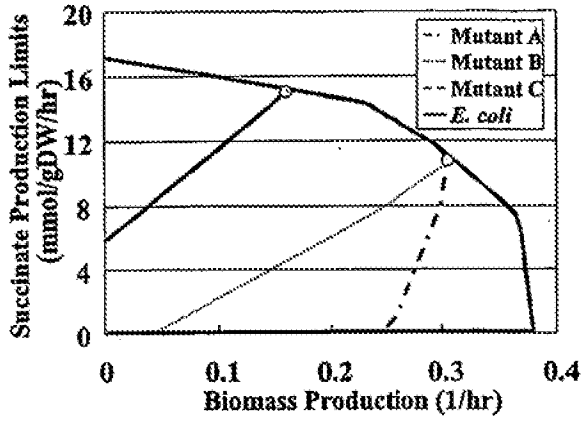


Figure 2

(A)


Succinate Production Limits (mmol/gDW/hr)
20
16
12
8
4
0
0
0.1
0.2
0.3
0.4
Biomass Production (1/hr)
Mutant A
Mutant B
Mutant C
E. coli

(B)

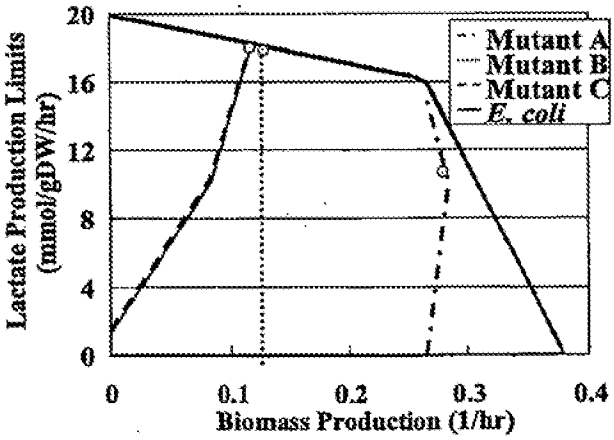
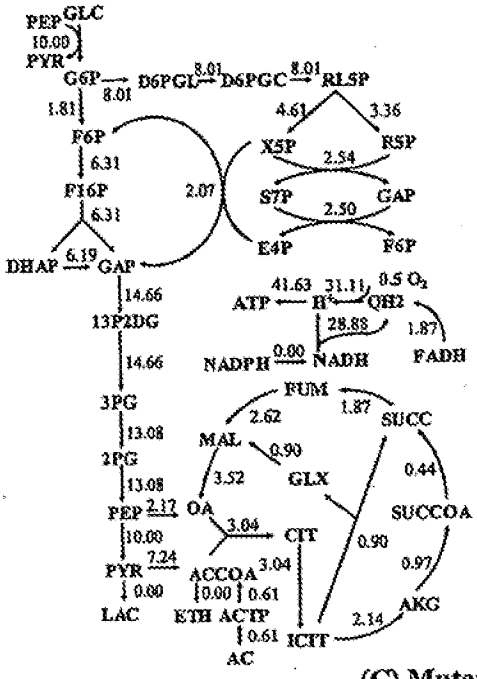
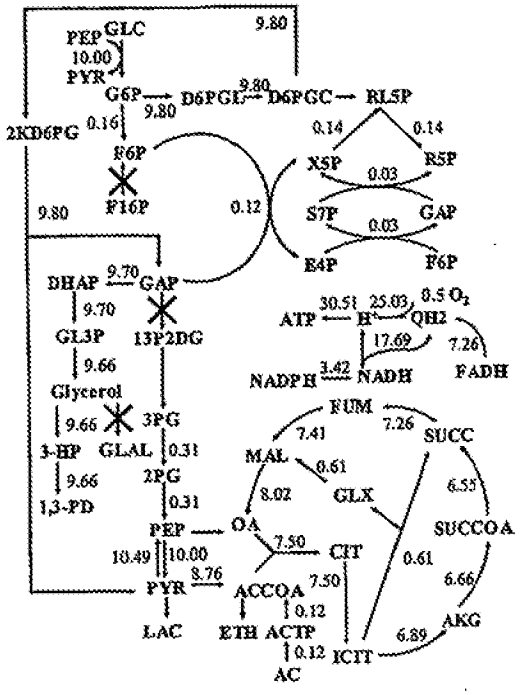

Lactate Production Limits (mmol/gDW/hr)
20
16
12
8
4
0
0
0.1
0.2
0.3
0.4
Biomass Production (1/hr)
Mutant A
Mutant B
Mutant C
E. coli

Figure 3

(A) "Wild-type" *E. coli*



(B) Mutant A



(C) Mutant B

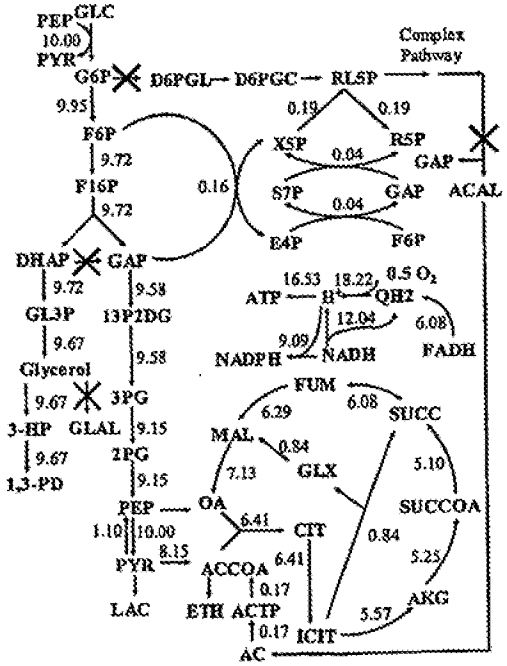


Figure 4

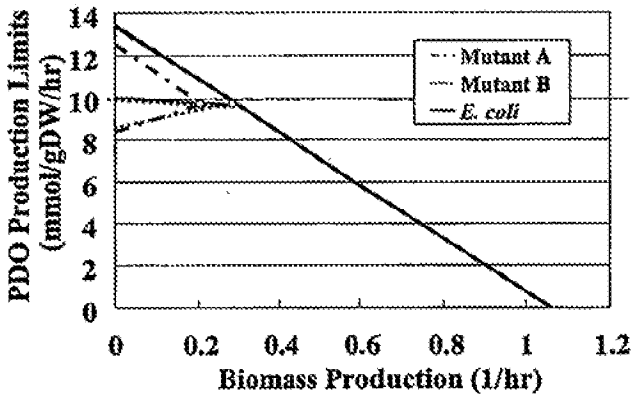

PDO Production Limits (mmol/gDW/hr)
14
12
10
8
6
4
2
0
0
0.2
0.4
0.6
0.8
1
1.2
Biomass Production (1/hr)
Mutant A
Mutant B
E. coli

Figure 5

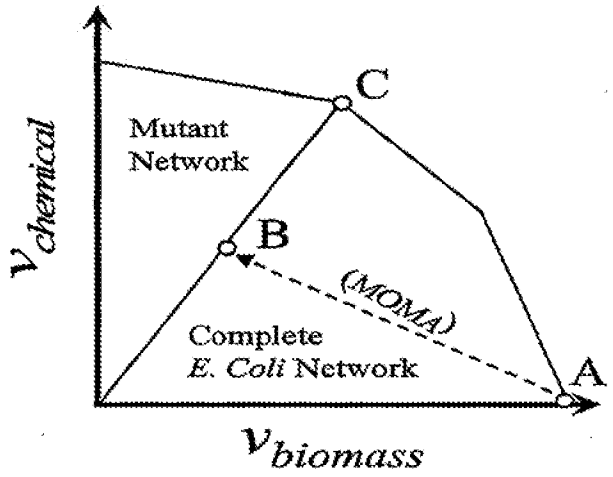

C
Mutant
Network
B
(MOMA)
Complete
E. Coli Network
A
$v_{chemical}$
$v_{biomass}$

Figure 6